Aboul-Ella Hassanien
Nilanjan Dey
Sally Elghamrawy   *Editors*

Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach

Springer

Studies in Big Data

Volume 78

The series "Studies in Big Data" (SBD) publishes new developments and advances in the various areas of Big Data- quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence including neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

** Indexing: The books of this series are submitted to ISI Web of Science, DBLP, Ulrichs, MathSciNet, Current Mathematical Publications, Mathematical Reviews, Zentralblatt Math: MetaPress and Springerlink.

More information about this series at http://www.springer.com/series/11970

Aboul-Ella Hassanien · Nilanjan Dey ·
Sally Elghamrawy

Editors

Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach

Springer

Editors
Aboul-Ella Hassanien
Department of Information Technology
Faculty of Computers and Artificial
Intelligence
Cairo University
Giza, Egypt

Nilanjan Dey
Department of Computer Science
and Engineering
JIS University
Kolkata, India

Sally Elghamrawy
Department of Computer Engineering
MISR Higher Institute for Engineering
and Technology
Mansoura, Egypt

Preface

In the fight against COVID-19 pandemic, big data analytics and artificial intelligence (AI) techniques play a significant role in several aspects. The integration between both technologies will help healthcare workers for early and accurately diagnosing COVID-19 cases. Both technologies allow the application of machine learning, deep learning, computer vision, and robotics to develop computer-based models for predicting, forecasting, diagnosing, and devolving drugs of the SARS-CoV-2 virus. In addition, the strategic planning for crisis management is supported by the aggregation of big data to be used in the epidemiologic directions. Moreover, AI and big data-driven tools present visualization for COVID-19 outbreak information that helps in detecting risk allocation and regional transmissions. This book presents different approaches for fighting against COVID-19 reflected in different categories, namely forecasting and visualization, diagnosis and predictions, artificial intelligence (AI), deep learning, and big data analytics. The content of this book is divided into four parts:

- *Forecasting and Visualization*
- *Diagnosis and Predictions of COVID-19*
- *Artificial Intelligence (AI) Against COVID-19*
- *Deep Learning Against COVID-19.*

Each submission is reviewed by the editorial board. Evaluation criteria include correctness, originality, technical strength, and significance, quality of presentation, and interest and relevance to the book scope. Chapters of this book provide a collection of high-quality research works that address broad challenges in both theoretical and application aspects of COVID-19 big data analytical and AI including machine learning and deep learning technologies.

Giza, Egypt Aboul-Ella Hassanien
Kolkata, India Nilanjan Dey
Mansoura, Egypt Sally Elghamrawy

Contents

About the Editors

Aboul-Ella Hassanien is the founder and head of the Egyptian Scientific Research Group (SRGE) and a professor of Information Technology at the Faculty of Computer and Artificial Intelligence, Cairo University. He has more than 1000 scientific research papers published in prestigious international journals and over 50 books covering such diverse topics as data mining, medical images, intelligent systems, social networks, and smart environment. He won several awards including the Best Researcher of the Youth Award of Astronomy and Geophysics of the National Research Institute, Academy of Scientific Research (Egypt, 1990). He was also granted a scientific excellence award in humanities from the University of Kuwait for the 2004 Award and received the superiority of scientific in technology—University Award (Cairo University, 2013). Also, he was honored in Egypt as the Best Researcher in Cairo University in 2013. He also received the Islamic Educational, Scientific and Cultural Organization (ISESCO) prize on Technology (2014) and received the State Award of Excellence in engineering sciences 2015. He holds the Medal of Sciences and Arts from the first class from president of Egypt in 2017.

Nilanjan Dey is an associate professor, Department of Computer Science and Engineering, JIS University, Kolkata, India. He is a visiting fellow of the University of Reading, UK. He was an honorary visiting scientist at Global Biomedical Technologies Inc., CA, USA (2012–2015). He was awarded his Ph.D. from Jadavpur University in 2015. He has authored/edited more than 70 books with Elsevier, Wiley, CRC Press, and Springer and published more than 300 papers. He is the editor-in-chief of the International Journal of Ambient Computing and Intelligence, IGI Global, and the associated editor of IEEE Access and International Journal of Information Technology, Springer. He is the series co-editor of Springer Tracts in Nature-Inspired Computing, Springer, series co-editor of Advances in Ubiquitous Sensing Applications for Health care, Elsevier, and series editor of Computational Intelligence in Engineering Problem Solving and Intelligent Signal Processing and Data Analysis, CRC. His main research interests include medical imaging, machine learning, computer-aided diagnosis, data mining, etc. He is the Indian ambassador of the International Federation for Information Processing—Young ICT Group and the senior member of IEEE.

Sally Elghamrawy is the head of Communications and Computer Engineering Department at MISR higher Institute for Engineering and Technology and a part-time associate professor at Electrical and Computers Engineering Department, Faculty of Engineering, British University in Egypt BUE and at Computers Engineering Department, Faculty of Engineering, Mansoura University in Egypt. She received a Ph.D. degree in 2012 in Distributed Decision Support Systems Based on Multi Intelligent Agents from Computer Engineering Department, Faculty of engineering, Mansoura University, and received a M.Sc. degree in Automatic Control Systems Engineering in 2006 from the same department, and received B.Sc. in Computers Engineering and Systems in 2003. She has been delivering lectures, supervising graduation projects, master's thesis, and doctoral dissertations. She was a practical trainer in the grants from the Ministry of Communications and Information Technology with

collaboration with IBM. She is certified in Cloud Application Developer 2018 Mastery Award for Educators, Mobile Application Developer Mastery Award for Educators 2017, and A+ International Inc. CompTIA. She is a member in Scientific Research Group in Egypt. Her research focuses on big data analysis, artificial intelligence, information retrievals, and software engineering. She is the author of number peer-reviewed publications, receiving best paper awards. She is also an IEEE member. She is an editor and reviewer in number of international journals and a judge on IEEE Young professionals' competitions.

Forecasting and Visualization

Coronavirus Spreading Forecasts Based on Susceptible-Infectious-Recovered and Linear Regression Model

Neha Tyagi, Naresh Dhull, Meenakshi Sharma, Vishal Jain, and Shashank Awasthi

Abstract An epidemiced environment is an occurrence of disease that spreads rapidly and disturbs many persons at the same time. Well, transmission of viruses is a big problem of today's era. Environmental conditions and habits of people are somehow responsible for the vulnerable conditions of the environment. In consequence, to find out the exact condition the study needs to know about the symptom of the person (like age, gender, current body temperature, dry cough duration, headache, and travel history). As per the study given by India today magazine, a very first case of coronavirus was found in Kerala's Thrissur district in India on January 30 2020. Thus, after 4–5 days, another two cases were found in Kerala too. Therefore, the state government took action and put these 3 patients in Isolation. Accordingly, with these 3 patients, the government quarantined their contacted persons (around 3400) too, who were suspected of symptoms of the coronavirus. Thereby, after the isolation and quarantine period, 3 patients among all were discharged upon recovery. Therefore, in this way, this study has to focus and trying to visualize the impact of isolation and quarantine on the health of the patients.

Keywords COVID-19 · Epidemic theory · SIR · Linear regression and coronavirus

N. Tyagi (✉) · S. Awasthi
G.L. Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: nehacs1988@gmail.com

S. Awasthi
e-mail: shashankglbitm@gmail.com

N. Dhull · M. Sharma
Galgotias University, Greater Noida, India

V. Jain
BharatiVidhyapeeth's Institute of Computer Applications and Management,
New Delhi, India

1   Introduction

The 2020 year, which unstable the lives on the earth, will be considered in the history of pandemics. It was first reported by the Wuhan Health commission, in China on 31st December 2019. From the 31st December 2019, it blowouts all over the world and damage the lives of male, female, kids, etc. The pandemic COVID-19 may be a major world health threat. The novel coronavirus have been rumored because of the most damaging metabolism virus since the 1918 contagion pandemic. Consistent with WHO novel coronavirus state of affairs report as on March 2020, a complete of around 600, more than 150 confirmed cases and more than 300 deaths are rumored across the planet [1]. Therefore, World unfolds has been fast, with More than 150 countries currently having rumored a minimum of one case. Coronavirus sickness (COVID-19) is an associate in nursing communicable disease caused by severe acute metabolism syndrome coronavirus-2. Coronavirus belongs to a family of viruses that is to blame for unhealthiness starting from respiratory disease to deadly diseases as MERS and SARS that were 1st discovered in China (2002) and Saudi Arabia (2012).

In view of this, the novel Coronavirus or higher referred to as COVID-19 was rumored in a metropolis, China for the 1st time on thirty-first Gregorian calendar month 2019. Consistent with Jiang et al. the death rate for this virus has been calculable to be 9/5 except for the people 80, this has gone up to 0.086 whereas for those >80 it's been noted to be 15% [2]. This has junction rectifier to aged persons higher than the age of 40 plus with underlying diseases like polygenic disorder, Parkinson's sickness, and disorder to be thought-about at the best risk. Symptoms of this sickness will take 5–20 days to look and might vary from fever, cough, and shortness of breath to respiratory illness, nephropathy, and even death [1]. The spread is one individual to another via metabolism droplets among shut contact with the typical variety of individuals infected by a patient being 4–5% however the virus isn't thought-about mobile [3]. Choice makers are benefited from a higher geared towards MC outputs complemented through min-max policies that foretell about the acute degrees of destiny possibilities with admiring to the epidemic [4].

Although, there exist an outsized variety of pieces of evidence wherever machine learning algorithms has proved to offer economical predictions in care [5–7]. Nsoesie et al. have provided a scientific review of approaches accustomed forecast the dynamics of contagion pandemic [8]. Similarly, they need reviewing analysis papers supported settled principle models, regression models, prediction rules, Bayesian network, SEIR model, ARIMA statement model, etc. Recent studies on COVID-19 embrace solely searching analysis of the offered restricted knowledge [9–11].

Here, the Prediction Module is projected for predicting the power of the patient to reply to treatment supported various factors e.g. age, infection stage, metabolic process failure, multi-organ failure, and therefore the treatment regimens. The Module implements the Whale improvement rule for choosing the foremost relevant patient options [12].

Thus, operative and effective vaccinum against novel coronavirus has not been unreal and thus a key half in managing this pandemic is to decrease the epidemic peak, conjointly referred to as flattening the epidemic curve. Here, the role of knowledge of scientists and data processing researchers is to integrate the connected data and technology to higher perceive the virus and its characteristics, which may facilitate in taking right selections and concrete arranges of actions. As a result in an even bigger image of taking aggressive measures in developing infrastructure, facilities, vaccines, and restraining similar to epidemic in the future. The objectives of this study area unit are following as designing a SIR (Susceptible, Infectious, Recovered) models to judge the unfold of sickness; Concluding the speed of unfolding of the sickness in the Republic of India &Forecasting of COVID-19 irruption exploitation SIR and Regression models.

Accordingly, the upcoming future is much unknown about this virus. The scientists, researchers, many peoples only can do predictions based on data given by the World health organization to the world.

The Introduction and Theory part given in Sect. 1, In order to this, the model used in this study illustrate in Sect. 2. In such a way, Analysis, performance evaluation and experimental results explicate in Sect. 3. Section 4 enlightens the Methodology discussion followed by the conclusion in Sect. 5.

2 Pandemic Theory

Koykul et al. examine about focal points and inconveniences of utilizing savvy lattice in circulated restoration vitality age [13]. Pipattanasomporn et al. talk about giving knowledge to a savvy network through a multi-operator framework [14]. The ideas remain identified with one another and bolster a ton of work just as parts. Generally the collaboration among supplier and performer is rapidly however here and there the change in encompassing may influence the communication of the supplier and the performer. Similarly, the plague hypothesis is likewise supportive to depict the idea of the spread of worm from source to a goal by characterizing the infection just as contortion or clamor in data. Likewise the survey study can say that the plague hypothesis depicts the proliferation of infection or mutilation or commotion in the data. Generally, the data consistently transmit or get regularly, so if the worm is engendered in a similar normal way, at that point the infection additionally proliferates similarly because the immediate contact characterizes the spread of infection using a typical frequency.

On condition that, there should be an occurrence of investigation of science, during the compelling contamination at the masses scale, two philosophies or ideas are portraying the engendering of disease which is characterized as the stochastic procedures and the deterministic techniques. Generally, a mass of n individuals is divided into a couple of compartments, and the spread of the illness is thought about (Fig. 1).
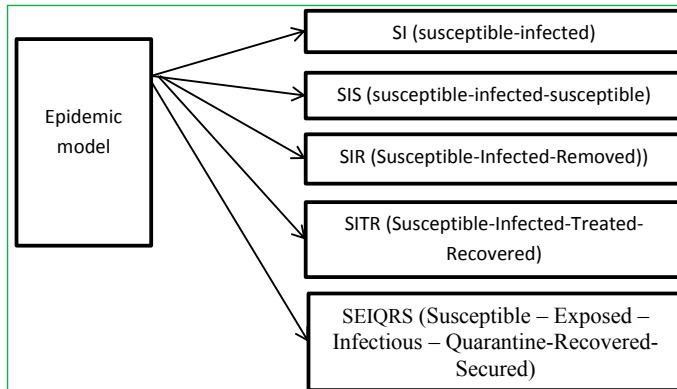
Fig. 1 Different types of epidemic models

3   SIR Model

The Kermack-McKendrick Model is very well known as the SIR model. Here for simulation, the study has used 3 variables, S–I–R to symbolize the total number of individuals in an epidemic environment at a given frame. Where susceptible individuals will depict by S, Infected individuals will depict by I and Recovered individuals or can say that the removed individuals from an epidemic environment that recovered by isolation or quarantine process by R.

Hence, these 3 kinds of persons may vary after some time frame. So, 't' will represent that time frame. The final number of functions of a specific time frame will be S (t); I (t) & R (t). In the same way, the SIR model is dynamic because of the number of people changing their state after the specified interval of time 't'.

On that account, to find out the value of conversion rate of I and S, then have to use the formula $\beta I/N$. β depicts the ordinary number of individuals contact per person per time that has to be multiplied by the disease spreading probability. Thus, the fraction of contact with individuals is depicted by I/N [15].

Where γ is the conversion rate of S and I i.e. rate of recovery. D is the duration of infection, at that juncture γ will depict by 1/D [15].

At this moment, a very important measurement of any disease prototypical is the R-Naught. It is the reproduction number of secondary contagions generated from one contagion individual in a total populace. As per the SIR model, $R0 = \beta N/\gamma$. The total population size will represent by $N = S+I + R$. In order to find out the value of β, then $\beta = R0\gamma/N$. R0 is property of breakout and does not change once calculated, as it's calculated assuming inclined inhabitants. The efficient copy quantity, "R" represents the typical collection of secondary infections in an inhabitant that can come with immune individuals. That is the metric this is wanted to be minimized via vaccination campaigns [15].

Thus, this is a very basic model but has effective assumptions. Very firstly, as a populace is shut and fixed, as such—nobody it included into the helpless gathering (means no births), all people that progress of from being contaminated to recuperated is for all time impervious to disease and there are no passings.

Similarly, as the populace considered as homogenous and just vary by their ailment state. At last third one, disease and that person's "infectiveness" or capacity to defect powerless people, happens at the same time [16, 17]. This section describes the simulating part of SIR model.

3.1 Simulation of SIR (the Susceptible-Infected-Removed) Model Using R Packages

As anticipated, the simulation study has used the deSolve R package for this simulation prediction. The study has used different variables here γ & β & t for several days. Here this research work had plotted the graph at 14 days for higher goals [15, 18].

Here, consider the R0 that is reproduction number. Because simulation found the data which is going to change day by day from COVID-19. As it a pandemic data onto the country, so, R0 represents the no. of cases produced by an infected person during the transmissible period. Thus, R0 is equal to cpd. So, with the help of R0, the value of β and exact information about the isolation period of the COVID-19 data can be calculated. For example, initially consider as the transmissible duration is equivalent to the period of contamination. It is notable that if one person is infected by a coronavirus, so that person will not infect the other persons immediately. It will take between 3 and 5 days as an incubation period and able to flacking the symptoms of coronavirus. For this reason, have to find out the value of β, i.e. $\beta = R0/\gamma$, where, R0 = 2.26 [15, 18].

Here, initially set the data onto 3 infected individuals. Then we will run R function of 14 days. The Study says that if R0 > one, the infection rate is greater than the recovery rate, and thus the infection will grow throughout the population. If R0 < one, the infection fast will die out since people are health-giving faster than they are transmitting it [15, 18].

Accordingly, Fig. 2 represents the stepwise step model for Susceptible-Infected-Removed, where, S (Susceptible), I (Infected), and R(No. of recovery people). Here this study is checking the top ten results come after applying the model on the discussed case. Findings have plotted the visualization by using the 't' parameter for 14 days [15, 18].

As a result, in Fig. 3, β gives the typical choice of contacts in line with the particular person in line with time multiplied by way of the likelihood of the disease spreading in a touch between I and S and γ is the conversion rate between I and S. Consequently, on the assumption of S = 0.9899906 I = 9.000750e−03 & R = 1.008659e−03, when β = 1.4247 and γ = 0.14286 [15, 18].

In order to this, Figs. 4 and 5, demonstrate the SIR model predictions in correspond

```
Model for Susceptible-Infected-Removed
SIR(14,0,0.32,1/7)
SIR<-function
(time,state,parameters)(with(as.list(c(state, parameters))
dS<- -(β*S*I)
dI<- ((β*S*I) – (γ*I))
dR<- γ*I
return list[dS,dR, dI]
β=1.427, γ=0.14286
```

Fig. 2 Model for susceptible-infected-removed using R packages
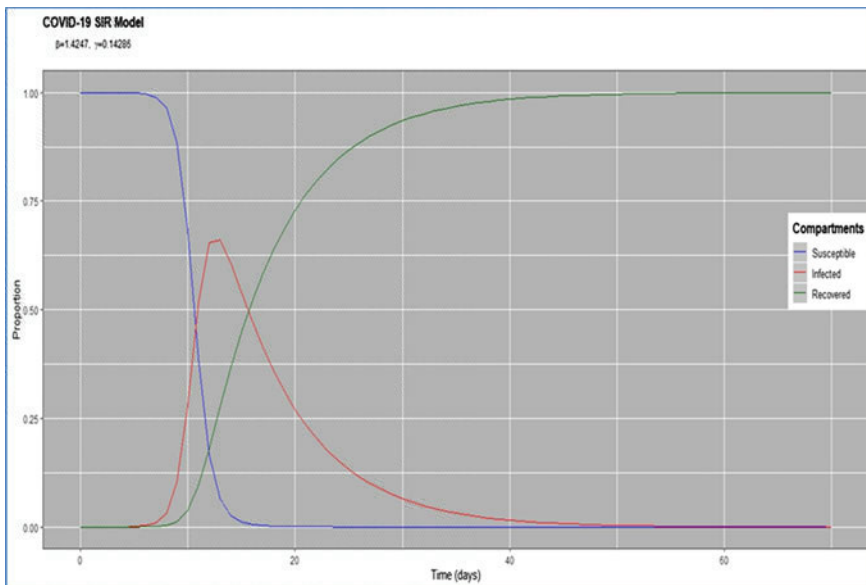
Fig. 3 Depicting the susceptible-infected-removed (SIR) model by using β and Susceptible (Blue), Infected (Red) and Removed (Green)

to the isolation period of the patients. This study also focuses on the isolation period which creates a great impact for the recovery of COVID-19 patients.
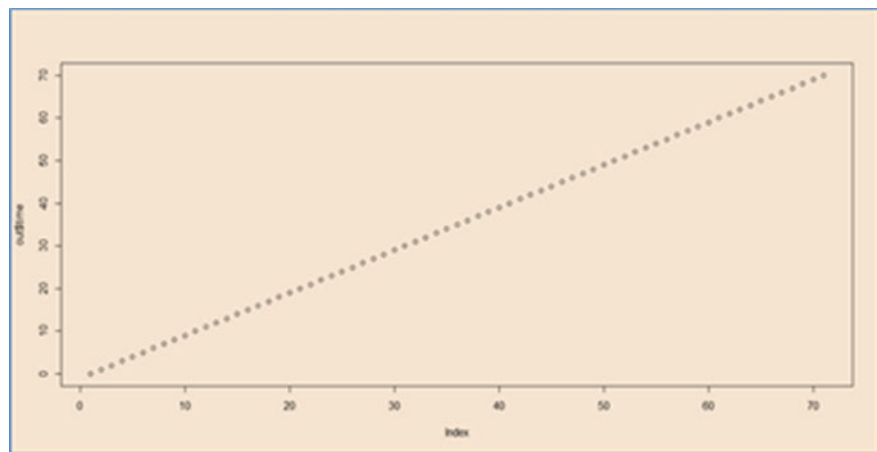
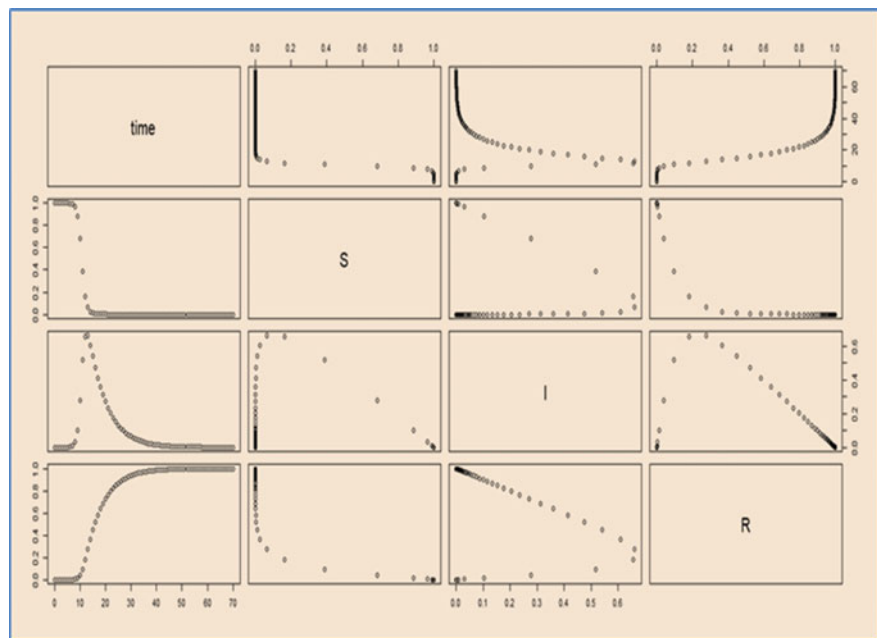**Fig. 4** SIR model in respect of time of isolation of populace

Fig. 5 SIR model representation in correspond to whole populace. The time span of simulation is 14 days

4 Linear Regression Model

Well, linear Regression approaches are numerical components of techniques that can approximation or are expecting a goal or variable that is dependent on the criteria of established variables. There are many types of Regression like Bayesian regression, polynomial regression, and linear regression, and so on. In this chapter regression type had used i.e. linear model for prediction of coronavirus pandemic. This model working is based on dependent and independent variables. The cost of slope and intercept is used to show the products of this model. Hence, in a linear regression model, $\beta 0$ and $\beta 1$ are used to signify an intercept & slope. Error rate will depict by $\epsilon. Y = \beta 0 + \beta 1x + \epsilon$ and $Y = \theta 0 + \theta 1x + \theta 2x2 + \theta 3x3 + \theta nxn$, [19]. For this purpose, in the given equations depicts the value of coefficients that assigned to predictors & polynomial diploma has to be depicted by n. Figure 8 depicts the Linear regression model fitting line. By using this plotted figure author says that as the time goes the number of confirmed cases of COVID-19 also goes high.

Accordingly, for proposing the model, here author considers the pandemic COVID-19 as a case study/baseline from India. Hence, more than half of the India has already stopped commuting. India has replaced handshakes with Namaste and now, the new standard: social distance. Around the early of March month, WHO has been stated a COVID-19 is a pandemic.

By using Fig. 6 author gives outline of the processes used to predict the epidemic environment data (COVID-19) in India. Here, the Impact Analysis to study the epidemiological environment, the author has used nCov2019 data outbreak. It includes real-time data and historical data as well. Initially these simulations are deploying the packages to start the prediction work. By virtue of basic functions are used for extracting the data example as: get_nCov2019 () which is used to get the latest online information; load_nCov2019 () which is used to get the historical data; summary which is used to access the whole data and plot is used to visualize the whole data in the pictorial form. Subsequently, the deployment process, now install all the packages by using commands: Remotes::install_github ("GuangchuangYu/nCov2019"); require (nCov2019); require (dplyr). Initial Impression: Here author have fixed two Variables A and B. A ← getnCov2019 () and B ← load_nCov2019 (). Following this it will give the total confirmed cases of the country. Initially this study explores the required data then after prediction functions has applied on particular data [3]. Similarly, Fig. 7, elaborate that how COVID-19 data gets trained for predictions. Author gives learning from initial to N number of steps to find the optimal prediction results (Fig. 8).

Fig. 6 Process followed for epidemic environment study
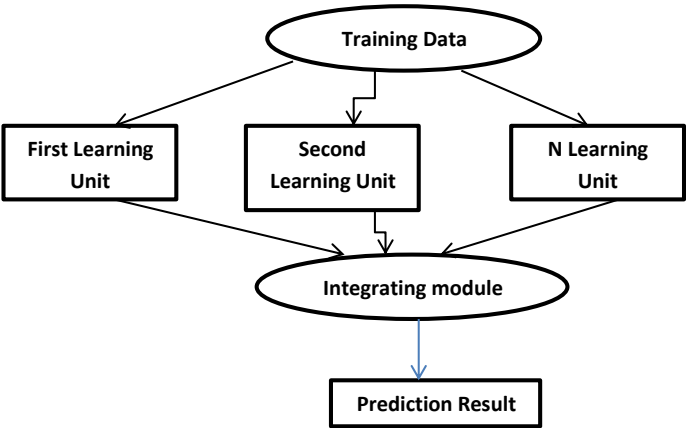
Fig. 7 Process to explore the COVID-19 training data
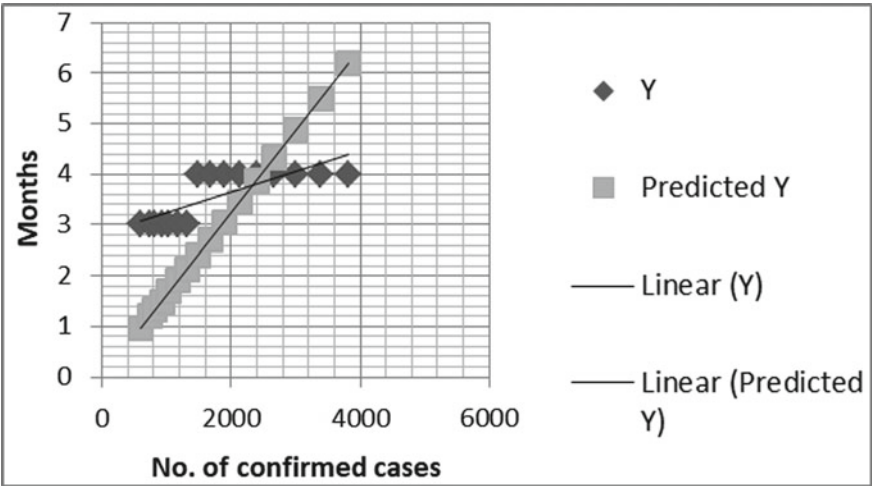
Fig. 8 Linear regression model fitting line

5 Comparison of SIR and Linear Regression Model

In Mentioned Table 1 is the Date-wise SIR, Linear Regression, and actual data comparison for 14 days; this chapter has used 14 days of training data onto prediction analysis. By using R packages for the SIR model depiction where with the help of this study found that $\beta = 1.4247$ and $\gamma = 0.14286$. Thereafter, Standard error in the linear regression model was 2.253 [17, 19]. The graphical assessment of Date-wise comparison of Susceptible-Infectious- Recovered, Linear Regression and actual data is given in Fig. 9. Subsequently, author finds out the linear regression model gives the

Table 1 Date-wise SIR, linear regression and actual data comparison for 14 days

Date	SIR	Linear	Actual
25/03/20	648	604	562
26/03/20	720	740	649
27/03/20	874	820	724
28/03/20	974	923	873
29/03/20	1081	1040	979
30/03/20	1199	1172	1071
31/03/20	1333	1321	1251
1/4/2020	1485	1488	1397
2/4/2020	1654	1676	1965
3/4/2020	1837	1887	2301
4/4/2020	2038	2127	2902
5/4/2020	2264	2396	3374
6/4/2020	2520	2669	4067
7/4/2020	2807	3007	4421
8/4/2020	3122	3386	5194
9/4/2020	3465	3814	5734
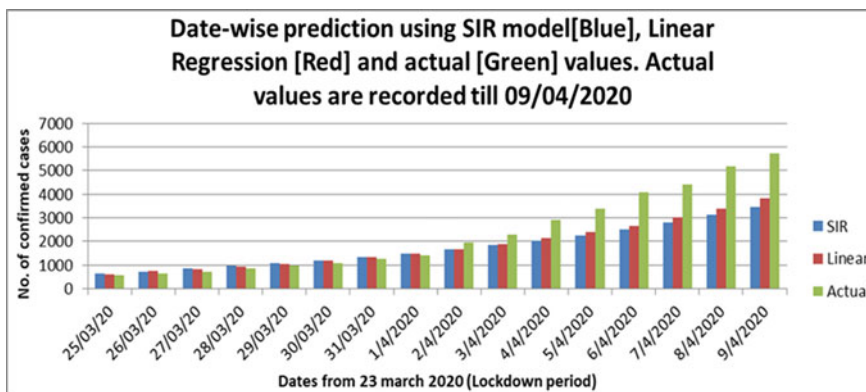
Fig. 9 Comparative graph of susceptible-infectious- recovered and linear regression model of COVID-19 Data

very close results to the actual one scenario. Similarly, Fig. 10, shows the COVID-19, actual data depiction in India before the period of lockdown and after the period of lockdown. So, it shows the change in the number of confirmed cases too.

Fig. 10 Change in number of confirmed cases before and after lockdown period due to COVID-19 in India

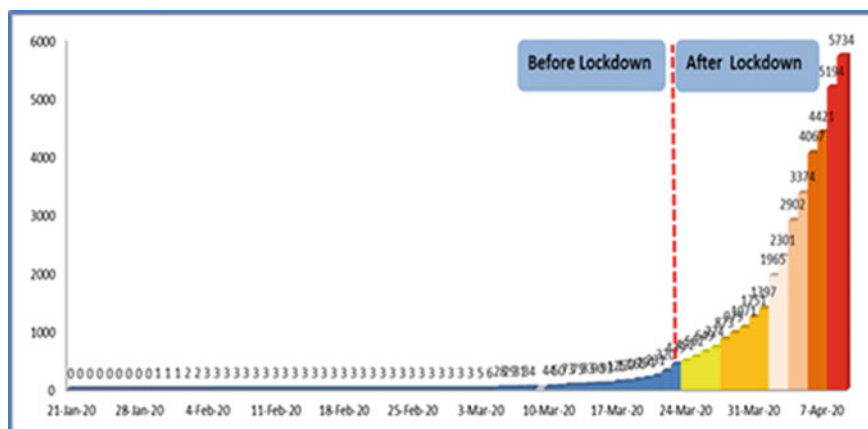## 6　Conclusion

This chapter focused on the COVID-19 initial case that occurs in India. Initially this chapter considered 3 cases of Kerala's district on January 30 2020. Thereupon, the government took steps to isolate or quarantine those 3 as well as the people those who had contact with those 3 individuals. Later, that 3 people have been recovered by the process of isolation or quarantine. As a consequence, by using the SIR model in the particular case mentioned by author, exploration work found that the recovery rate has grown by using the isolation process for infected people. Initially set the value of R is 0.000000e+00 after the SIR model application will find R = 1.008659e−03. Therefore, it shows the Isolation and quarantine help to improve the health condition of COVID-19 infected people. Similarly, the problem has solved by using a linear regression model. Simulation results shows that the standard deviation between SIR and actual is ±15% and the standard deviation between linear regression and actual is ±7%. Thus, it has been concluded that linear regression model predicts the very near results of the actual data.

References

1. World Health Organization.: Coronavirus disease 2019 (COVID-19): situation report, **67** (2020)
2. Jiang, F., Deng, L., Zhang, L., Cai, Y., Cheung, C.W., Xia, Z.: Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). J. General Int. Med. 1–5 (2020)
3. Wu, Z., McGoogan, J.M.: Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA (2020)

4. Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. Appl. Soft Comput. 106282 (2020)

5. Ye, Q.H., Qin, L.X., Forgues, M., He, P., Kim, J.W., Peng, A.C., Simon, R., Li, Y., Robles, A.I., Chen, Y., Ma, Z.C.: Predicting hepatitis B virus–positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. Nat. Med. **9**(4), 416–423 (2003)

6. Mai, M.V., Krauthammer, M.: Controlling testing volume for respiratory viruses using machine learning and text mining. In: AMIA Annual Symposium Proceedings, vol. 2016, p. 1910. American Medical Informatics Association

7. Purcaro, G., Rees, C.A., Wieland-Alter, W.F., Schneider, M.J., Wang, X., Stefanuto, P.H., Wright, P.F., Enelow, R.I., Hill, J.E.: Volatile fingerprinting of human respiratory viruses from cell culture. J. Breath Res. **12**(2), 026015 (2018)

8. Nsoesie, E.O., Brownstein, J.S., Ramakrishnan, N., Marathe, M.V.: A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza Other Respir. Viruses **8**(3), 309–316 (2014)

9. Pirouz, B., ShaffieeHaghshenas, S., ShaffieeHaghshenas, S., Piro, P.: Investigating a serious challenge in the sustainable development process: analysis of confirmed cases of COVID-19 (New Type of Coronavirus) through a binary classification using artificial intelligence and regression analysis. Sustainability **12**(6), 2427 (2020)

10. More, G.D., Dunowska, M., Acke, E., Cave, N.J.: A serological survey of canine respiratory coronavirus in New Zealand. New Zealand Vet. J. **68**(1), 54–59 (2020)

11. Wu, C., Chen, X., Cai, Y., Zhou, X., Xu, S., Huang, H., Zhang, L., Zhou, X., Du, C., Zhang, Y., Song, J.: Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan. China, JAMA Intern. Med. (2020)

12. Diagnosis and Prediction model for COVID19 patients response to treatment based on convolutional neural networks and whale optimization algorithm using CT image. medRxiv 2020.04.16.20063990; https://doi.org/10.1101/2020.04.16.20063990

13. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Comput. Netw. **38**(4), 393–422 (2002)

14. Vahdat, A., Becker, D.: Epidemic routing for partially connected ad hoc networks. Duke University Technical Report CS-200006 (2000)

15. https://cran.rproject.org/web/packages/shinySIR/vignettes/Vignette.html

16. CSSEGISandData. CSSEGISandData/COVID-19 (2020) https://github.com/CSSEGISandData/COVID-19

17. Pandey, G., Chaudharya, P., Guptab, R., Pal, S.: SEIR and regression model based COVID-19 outbreak predictions in India (2020)

18. https://rpubs.com/choisy/sir

19. Gupta, R., Pal, S.K.: Trend analysis and forecasting of COVID-19 outbreak in India. Published online 30/03/2020 in preprint archive MedRxiv https://www.medrxiv.org/content/10.1101/2020.03.26.20044511v1 (2020)

Virus Graph and COVID-19 Pandemic: A Graph Theory Approach

H. R. Bhapkar, Parikshit N. Mahalle, and Prashant S. Dhotre

Abstract In the field of science and technology, the graph theory has offered several approaches to articulate any situation or concept. The use of graph theory enables the users to understand and visualize the situations like COVID-19. Looking at this pandemic disease, its impact and the preventing measures, the graph theory would be the most appropriate way to exercise the graph models with theoretical as well as practical aspects to control this epidemic. In the context of COVID-19, this chapter defines the variable set, variable graphs, and their types considering the variations in the vertex sets and edge sets. The virus graph and their type are discussed in this chapter that states that the Virus graph type I and III are not so perilous for all living beings, but virus graph type III and IV are extremely hazardous for the harmony of the world. Initially, the COVID-19 was in Virus graph-I type, but presently it is in Virus graph-II type. Given different aspects for expansion of pandemic, this chapter presents growth types of virus graphs and their variation as 1-1, 1-P, and 1-all growth types. This chapter provide the number of infected people after 'n' number of days concerning different values of P and growth rates with $I0 = 100$. At the end of this chapter, the country-wise starting dates of stages of the virus graph-I and II are specified. The concept of cut sets is applicable for the prevention of COVID-19 and the whole world is using the same analogy.

H. R. Bhapkar (✉)
Department of Mathematics, MIT ADT University's, MIT School of Engineering, Pune, Maharashtra, India
e-mail: hrbhapkar@gmail.com

P. N. Mahalle
Department of Computer Engineering, STES'S, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India
e-mail: aalborg.pnm@gmail.com

P. S. Dhotre
Department of Information Technology, JSPM's Rajarshi Shahu College of Engineering, Pune, Maharashtra, India
e-mail: prashantsdhotre@gmail.com

1 Introduction

COVID-19 is the transferrable disease caused by the recent coronavirus recently started in Wuhan, China. This virus and subsequently the disease were shadowy to the world before its outbreak. Considering the recent COVID-19 virus and its spread across the globe, it is important to understand and visualize the virus spread and impact. The disease caused by this virus has become pandemic and many countries are affected badly. Using the graph theory approach, this chapter helps users to understand and visualize this disease, impact and spread. The different graph method presented in this chapter shows a virus, its growth type is presented using graph theory. The number of persons who are affected and prevention is also presented in this chapter. This chapter concludes that there is an infinite scope of mathematics for the research as well as resolving social problems like COVID-19 and technical problems.

The reader will refer [1–4] for the absolute dealing with the subject matter. All Graphs considered in this chapter are simple as well as connected. The neighbor of the vertex v in graph H is the set of all the vertices adjacent to the vertex v in H. A graph with n vertices and without any edges is called the Null graph and it is denoted by Nn [1, 5–9].

A simple connected graph, in which the degree of each vertex is 2, is called a cycle graph. Cn is the cycle graph on n vertices [4]. A graph, in which one vertex is adjacent to n pendant vertices, is called the star graph. It is symbolized by K1, n. Here, $|V(K1, n)| = n + 1$ along with $|E(K1, n)| = n$ [10].

The chapter is organized as follows. Basic terms of graph are presented in Sect. 1. Motivation and related work is presented in Sect. 2 and 3 respectively. Graphical theoretical model that emphasizes on Virus Graph I, II, III and IV is presented in Sect. 4. Sections 5 and 6 discusses on growth rate and its types. Country wise stages of Virus graph I and II is presented in Sect. 7. Growth rate of COVID-19 is predicted and presented in Sect. 8. This chapter is summarized with future outlook in Sect. 9 [10].

2 Motivation

Almost 2.5 million cases of COVID-19 (corona virus) and more than 160,000 deaths have now been reported worldwide [4]. The largest part of the epidemic in the world comes into sight to be steady or declining [3]. A good number countries are immobile in the early stages of their epidemics and few of them were affected early in the pandemic are now starting to see an improvement in cases. Hence, it is the most important and essential to prevent the spread of such types of epidemic [10]. As we

know, mathematical modelling award different astonishing inspiration and tools to study different communal as well as technical problems and interpret solutions. This will lead to find the practical solutions of a variety of problems and helps to continue the harmony of the mankind.

3   Related Work

Let us consider following terminologies.

- **Partite Graphs**

A graph H is n- partite graph if $V(H) = V_1 \cup V_2 \cup \ldots, \cup V_n$, where all V_i are disjoint and every edge of H joins a vertex of V_i and V_j for $i \neq j$. If n = 2, 3, 4 then graphs are called Bipartite, Tripartite and four partite respectively [1].

- **Cut Sets**

A set of edges of a connected graph H, whose removal disconnects H, is called the disconnecting set of H. The smallest disconnecting set is called the cut set of H [2, 4].

- **Corona Product of Graphs**

The corona product of H and K is denoted by $H \circ K$ and obtained from a copy of H and | V (H) | copies of K, joining each vertex of H to all vertices of the graph K. This graph product was initiated by mathematicians Frucht and Harary in 1970 [11, 12].

Giulia Giordano et al. [2] proposed an innovative epidemic model which distinguishes between infected people depending on whether they have been confirmed by considering their symptoms. Non diagnosed peoples can spread virus rapidly than diagnosed people. Therefore the divergence between diagnosed and non-diagnosed plays an important role for the deterrence of a pandemic. A fuzzy theory and deep learning networks help to enhance for acquiring superior stochastic insights concerning the epidemic growth is experimented [3]. A deep learning based Composite Monte-Carlo (CMC) showed better results than simple Monte Carlo (MC) which will be obliged for decision makers for greater ranges of the future promises of epidemic and pandemic [13]. In [14], it is studied that due less and incorrect information about COVID-19, there is no any exact model which can predict spread of the pandemic. Every model has different levels of predictive efficiency.

By analyzing data in few countries, it is noted that an infection reached to peak around 10 days after the controlling measures are initiated. The growth rate of infected people was slowly decreasing during this period. But, especially the growth rate in Italy remains exponential. Hence, quarantine is insufficient and need strict measures [15]. In [16], the authors have made a mathematical model for the epidemic by applying linear differential equations. By identifying patterns and analyzing desired data, it is concluded that the growth rate is dynamic or exponential depending upon precautions taken by people. Shinde Gitanjali et al. [17], presented and meticulously

discussed different predictive analytic models as well as algorithms for the number infected cases in the near future. Moreover, the Prophet predictive analytics algorithm is implemented on the Kaggle dataset and its predictions are studied in their research work. The following new terms are defined for constructing the pedestal of mathematical modelling of any types of pandemic or COVID-19.

Critical decision making is difficult due to uncertainty caused by novel coronavirus epidemic. Fong et al. [10] presented deep learning and fuzzy based prediction method for the future possibilities of coronavirus and its impact. The present events and its future behavior is presented using Composite Monte Carlo simulation method. The difficult task is accurate forecasting of destiny of an epidemic is presented by Fong et al. [18] using augmentation of existing data, panel design for selection of best forecasting model and its fine tuning of parameters of each model. Deep learning method was presented by Hu et al. [19] for forecasting of COVID-19.

Based on the lung CT scan images, Rajinikanth [20] presented a system for detection of COVID-19. This proposed method is based on Otsu and a meta-heuristic Harmony search algorithm. Using graph theory, data classification was proposed by Kamal [21] which is based on De-Bruijn graph with MapReduce framework.

After evaluation of related work, there is a need for the mathematical modelling and visualization of COVID-19 using graph theory is essential to spread awareness among many stakeholders.

3.1 Variable Set

A set S is said to be Variable set if elements of the set S changes with respect to time or some rule. That is, the set S is not constant set. Its cardinality changes with respect to time. Sv is the notation of variable set. In the variable sets, time units depend upon its nature. According to the scenery of the cardinality of the set S, there are three types of sets.

- Increasing Variable Set: A variable set Sv is said to be increasing variable set if $|S_v(x)| < |S_v(y)|$, whenever $x < y$, where x and y are different times.
- Decreasing Variable Set: A variable set S_v is said to be decreasing variable set if $|S_v(x)| > |S_v(y)|$, whenever $x < y$.
- Non Decreasing Variable Set: A variable set S_v is said to be non-decreasing variable set if $|S_v(x)| \leq |S_v(y)|$, whenever $x \leq y$.
- Non Increasing Variable Set: A variable set S_v is said to be non-increasing variable set if $|S_v(x)| \geq |S_v(y)|$, whenever $x \leq y$.
- Stable Variable Set: A variable set S_v is said to be stable variable set if $|S_v(t)| =$ constant, for any time t. However, the set is a variable set. Elements of the set S vary according to time, but the $|S_v(t)|$ is steady, for any time t.

3.2 Variable Graph

A graph H is said to be a vertex variable graph if V (H) or E (H) is variable sets. Variable graphs are also known as V-graphs. Big network graphs are variable graph. There are two types of variable graphs.

3.3 Edge V-Graph

A variable graph H is said to be edge V-graph if E (H) is a variable set and V (H) is the stable variable set.

3.4 Vertex V-Graph

A variable graph H is said to be vertex V-graph if V (H) is a variable set and E (H) is the constant variable set.

3.5 N-Partite V-Graphs

A variable graph H is said to be n-partite V-Graph if

1. V (H) $= V_1$ U V_2 U V_3,….,V_n where V_1, V_2,V_3,…, V_n disjoint variable sets having different characteristics.
2. There exists a bond on the link or edge between vertices of Vi & Vj, for i, j and i $\neq$ j.

3.6 Bipartite V-Graph

A variable graph H is said to be Bipartite V-Graph if

1. V (H) $= V_1$ U V_2, where V_1 and V_2 are disjoint variable sets with different characteristics.
2. There exists a bond on the link or edge between vertices of V_1 and vertices of V_2
3. There is no any bond among the vertices of V_1 only or V_2 only.

These types of graphs are denoted by BV_2. In BV_2, a vertex x of V_1 is said to be **Active Vertex** or element if there exists a bond between x and at least one vertex of V_2 or x is trying to build a bond or edges to the vertices of V_2. Moreover, x is ready for the sharing some characteristics. Other vertices of V_1 are known as the Passive
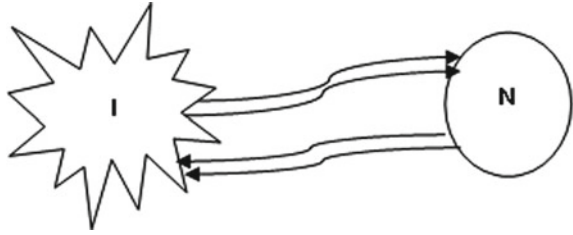
Vertices. A vertex y of V_2 is said to be Active Vertex or element if there exists a bond between y and at least one vertex of V_1 or y is aiming to build a bond or edges to the vertices of V_1. Other vertices of V_2 are known as the **Passive Vertices**.

4   Graph Theoretical Model

4.1   *Virus Graph I*

A Bipartite V-graph H is said to be Virus Graph I (VRG-I) if

1. $V(H) = I \cup N$, where, I be the variable set of vertices have some special properties or infected by virus and N be the variable set of vertices not having a virus.
2. If $x \in I$, creates a bond or an edge with the vertex $y \in N$ or vice versa, then y is shifted to I and $N = N - \{y\}$.
3. If $x \in I$, is recovered by treatment or lost properties of virus then, x is shifted to N and $N = N \cup \{x\}$. The diagrammatic representation of VRG-I is shown in Fig. 1.

4.2   *Virus Graph II*

A Tripartite V-graph H is said to be Virus Graph II (VRG-II) if

1. $V(H) = I \cup N \cup F$, where, I be the variable set of vertices having the virus, N be the variable set of vertices not having virus and F be the set of vertices which can never be shifted to I or N.
2. If $x \in I$, creates a bond or an edge with the vertex $y \in N$ or vice versa, then y is shifted to I and $N = N - \{y\}$.
3. If $x \in I$, is recovered by treatment or vanished properties of virus then, x is shifted to N and $N = N \cup \{x\}$.
4. Vertices of I are shifted to F if the vertices are infected forever. Therefore, S is the non-decreasing variable set. This is represented in Fig. 2.